

Sugerindo novas conexões entre ruas da rede urbana baseado em cidades planejadas

Daniel Souza de Campos¹

¹Departamento de Ciência da Computação - Universidade Federal de Minas Gerais (UFMG)

daniell1@ufmg.br

Abstract. *The expected population increase for the next few decades makes cities have to prepare to accommodate all new people. The best paradigm to follow to solve this problem is still up for debate. However, an apparent consensus is that planned cities can do better. The layout of street networks and what makes a city considered planned can be analyzed using metrics applied to the graph in which the city was modeled. This work augments a representative dataset of global cities with new metrics in order to filter out those cities considered planned. These cities will serve as input for a future Graph Machine Learning model to suggest new streets in cities not considered planned.*

Resumo. *O aumento populacional esperado para as próximas décadas faz com que as cidades tenham que se preparar para acomodar todas as novas pessoas. O melhor paradigma a seguir para resolver tal problema ainda é um debate. Entretanto, um aparente consenso é que cidades planejadas podem se sair melhor. A disposição da rede de rua e o que faz uma cidade ser considerada planejada podem ser analisadas por meio de métricas aplicadas ao grafo na qual a cidade foi modelada. Este trabalho aumenta um dataset representativo de cidades globais com novas métricas a fim de filtrar aquelas cidades consideradas planejadas. Essas cidades servirão de entrada para um modelo de Graph Machine Learning futuro para sugerir novas ruas em cidades não consideradas planejadas.*

1. Introdução

Redes de ruas são críticas para o transporte local e nacional. Ter uma infraestrutura bem planejada é um grande diferencial para a economia e bem estar geral da população de uma cidade[Burghardt et al. 2022]. Pode-se dizer que a disposição atual das ruas é um reflexo de padrões de desenvolvimento fomentados por necessidades históricas, limitações geográficas e políticas de expansão [Burghardt et al. 2022][Uhl et al. 2022].

Existem vários estudos com relação a organização das ruas de uma cidade. Alguns focam na análise da expansão da malha urbana ao longo dos anos até onde a existência de dados permite [Burghardt et al. 2022] [Barrington-Leigh and Millard-Ball 2020] e outros tentam resgatar como se deu essa expansão ao completar dados e comparar mapas de várias épocas [Uhl et al. 2022]. Diversos trabalhos tentam identificar quais são as principais vias de uma cidade do ponto de vista da teoria de redes complexas [Cao et al. 2016], transporte de informações[Scellato et al. 2006], robustez [Masucci and Molinero 2016], vulnerabilidade [Jenelius and Mattsson 2015][Li and Liu 2020] e trafegabilidade[Hu et al. 2008].

A expansão da rede de ruas por meio de disposições em grade (*grid*), ou seja, cruzamentos entre 4 vias, aumenta a trafegabilidade, mobilidade e acessibilidade de pedestres [Hu et al. 2008] e é característica de cidades bem planejadas [Rui et al. 2013]. Entretanto, nas últimas décadas, foi observado um movimento de expansão que tendeu a utilizar de ruas mais longas e grupos menos conectados como, por exemplo, no caso de condomínios fechados, que aumentam a necessidade de uso de automóveis particulares [Barrington-Leigh and Millard-Ball 2020].

Recentemente surgiu o interesse da comunidade na aplicação e desenvolvimento de modelos de *Machine Learning* para ajudar a entender peculiaridades da estrutura de uma cidade. Um exemplo é a estimação de velocidade máxima de vias [Jepsen et al. 2022] [Jepsen et al. 2018]. Esses trabalhos que envolvem *Machine Learning* e outros estudos comumente modelam a rede de ruas como um grafo do tipo primal onde os seus vértices são os cruzamentos e as arestas são as ruas que se conectam aos cruzamentos [Jepsen et al. 2022] [Jepsen et al. 2018] [Jenelius and Mattsson 2015] [Cao et al. 2016].

Dessa forma, dada a importância geral da rede de ruas para cidades, deve-se pensar em maneiras de melhorar a sua disposição por meio de futuras intervenções com o auxílio de ferramentas de sugestão de novas conexões entre as ruas.

A proposta geral do trabalho a ser alcançado na POC2 é desenvolver um modelo capaz de prever e sugerir novas conexões entre ruas. Além disso, o modelo deve levar em consideração a estrutura inerente de cidades consideradas planejadas para que a organização viária urbana de outras cidades possam melhor se aproximar dessas configurações. O modelo desenvolvido poderá ser completamente novo ou baseado em modelos de *Machine Learning* para grafos já existentes. Entretanto, o objetivo específico da POC1 é o de conseguir, primeiramente, um conjunto de cidades consideradas planejadas com base em métricas que possam vir a serem utilizadas como dados de treino no futuro modelo de *Graph Machine Learning*.

2. Metodologia

A seguir é apresentada a metodologia adotada que representa as três tarefas principais cumpridas ao longo deste trabalho.

- **Pesquisa Literária:** Realizar pesquisa e estudo de trabalhos sobre a modelagem, análise e expansão da rede urbana. Além disso, encontrar métricas utilizadas para a classificação qualitativa e quantitativa de cidades sob aspectos de interesse como conectividade e eficiência.
- **Aquisição de dados:** Pesquisar fontes de dados sobre redes urbanas que possam ser utilizadas para a aplicação do futuro modelo.
- **Filtro de dados:** De acordo com as métricas encontradas e os dados disponíveis, separar um conjunto de cidades aptas a serem utilizadas como dados de treino de melhor qualidade para o modelo de *Graph Machine Learning* a ser desenvolvido na POC2.

3. Atividades Realizadas por Semana

Segue a lista de atividades realizadas semanalmente durante este período da POC1. Essa lista não seguiu o cronograma da Proposta Inicial porque se tornou clara, após a

apresentação parcial, a necessidade de um maior entendimento do problema antes de focar no estudo de *GraphML*. Dessa forma, estudos relacionados a *Machine Learning* foram deixados para o período da POC2. Entretanto, pode-se dizer que as tarefas realizadas ainda estão em conformidade geral com a metodologia pretendida.

Semana	Tarefa
21/09	Estudo de modelos de GML e trabalhos sobre disposição de ruas
28/09	Pesquisa de recursos para baixar redes de ruas como grafos
05/10	Web Scrape de lista de cidades planejadas no Wikipedia
12/10	Coleta dos grafos de cidades planejadas do Wikipedia
19/10	Coleta dos grafos de cidades planejadas do Wikipedia
26/10	Apresentação Parcial e Estudo da teoria de expansão de cidades
02/11	Estudo da teoria de expansão de cidades
09/11	Identificação de métricas e suas aplicações na rede urbana
16/11	Identificação de métricas e suas aplicações na rede urbana
23/11	Identificação de métricas e suas aplicações na rede urbana
30/11	Download de conjunto de dados encontrado e computação de novas <i>features</i>
07/12	Computação de novas features e análise dos resultados
14/12	Análise dos resultados e definindo escolha de cidades planejadas

4. Referencial Teórico

4.1. Grafos e a modelagem de cidades

A rede de ruas de uma cidade pode ser naturalmente modelada como um grafo. Grafos são formulações matemáticas utilizados para expressar relações entre um conjunto de objetos. Formalmente, um grafo G é constituído por um conjunto de vértices V , também conhecidos como nós, e outro de arestas E , direcionadas ou não, que os conectam entre si formando, então $G(V, E)$ [Wilson 1996].

Existem diferentes abordagens para representar o espaço urbano como um grafo, cada uma com suas características e particularidades. A mais utilizada é a abordagem do grafo primal. Essa abordagem modela a rede viária de forma que cruzamentos sejam vistos como nós e ruas como arestas entre os nós. Essa definição reinterpreta o espaço urbano preservando a correspondência entre as dimensões das entidades físicas e as dimensões das entidades abstratas. Dessa forma, entidades zero-dimensionais tal como interseções entre ruas são interpretados como entidades zero-dimensionais no grafo, ou seja, nós. Entidades unidimensionais como ruas são interpretadas como entidades unidimensionais no grafo, ou seja, arestas [Agryzkov et al. 2017].

Uma outra modelagem poderia ser: ruas como nós e, para toda outra rua na qual uma rua possui uma junção, existe uma aresta as conectando. Essa abordagem é conhecida como o modelo dual. A dualidade, em termos gerais, pode ser definida como uma translação de um conceito em outro de maneira um para um por uma operação de involução. Dessa forma, leva-se as entidades unidimensionais do grafo primal para as zero-dimensionais no dual e vice-versa. Dessa forma, nós do grafo primal se tornam arestas no dual e arestas no primal se tornam nós no dual [Agryzkov et al. 2017].

O grafo dual apresenta diversas variações sobre a abordagem seguida para representar os nós do grafo primal em arestas no dual. Esse processo é chamado de modelo de generalização [Porta et al. 2005].

Um primeiro modelo é conhecido como a abordagem da rua nomeada [Porta et al. 2004] [Agryzkov et al. 2017]. Nesse método, o princípio de continuidade é o nome da rua, ou seja, arestas que estejam associadas ao mesmo nome de rua serão condensadas em apenas um nó no grafo dual. O principal problema dessa abordagem é que ele requer que todas as arestas no grafo primal possuam um nome associado. Isso está sujeito a erros e a disponibilidade de tal informação pode ser precária. Além disso, nomes de ruas não tem qualquer significado topológico ou espacial, são apenas uma nomenclatura [Porta et al. 2004].

Outras três abordagens são: Linearidade do espaço de ruas em que une-se arestas que possuam um ângulo obtuso entre elas [Agryzkov et al. 2017]; tradução de faces do grafo primal em nós no dual e arestas unindo faces no grafo primal se tornam arestas no dual [Agryzkov et al. 2017] e a Negociação de Continuidade de Intersecção (ICN) [Porta et al. 2004] onde, iterativamente, são unidas arestas com base no ângulo convexo entre elas.

Outro tipo de grafo dual é aquele definido pela área da Sintaxe Espacial (*Spacial Syntax*). Essa área cria o chamado Mapa Axial onde cada espaço reto é representado por uma única linha reta, a linha axial. Depois, cada linha axial é transformada em um nó enquanto cada intersecção entre quaisquer pares de linhas axiais é transformada em uma aresta. Depois, medidas de acessibilidade (integração) são calculadas sobre esse grafo chamado de grafo de conectividade [Porta et al. 2004].

Existem vantagens e desvantagens em escolher o modelo dual sobre o primal. A principal desvantagem do modelo dual é que, no seu processo de construção, a ideia de distância é perdida. Assim, quão maior for o comprimento de uma rua, maior o seu grau provável no grafo dual. Dessa forma, a representação dual associada com um modelo de generalização distancia o sistema de ruas urbano do domínio geográfico [Porta et al. 2005]. Entretanto, a sua desvantagem também é dita como uma vantagem pois não é necessário se preocupar com distâncias, apenas conexões e agrupamentos entre ruas [Porta et al. 2004].

A teoria de redes complexas define algumas redes como sendo sem escala (*scale free*). É característica desse tipo de rede que a distribuição do grau dos nós siga a chamada *power-law* da forma: $p_k \propto \frac{1}{k^\alpha}$, onde k é o grau do nó e α é uma constante. Assim, existem muito mais nós de grau baixo do que alto. As redes de ruas vistas sob o olhar dual podem ser consideradas *scale-free* pois, ao agregar arestas do grafo primal em apenas um nó e dado que existem poucas grandes avenidas e muitas ruas locais, existirão muitos nós de grau baixo e poucos de grau alto [Porta et al. 2004] [Xie and Levinson 2007].

Outra categoria de redes são as chamadas aleatórias. Nesse tipo de redes, espera-se que a distribuição de graus siga uma distribuição de Poisson centrada em um valor λ . Nesse caso, a maioria dos nós possuem pouca variação na quantidade de conexões. A abordagem primal da rede de ruas pode ser considerada aleatória já que a maioria das intersecções são entre 3 ou 4 vias [Porta et al. 2004] [Xie and Levinson 2007].

4.2. Graph Machine Learning

Existem modelos de *Machine Learning* que operam sobre os grafos utilizando das possíveis *features* que acompanham os seus componentes. Essa sub-área de *Machine Learning* é conhecida como *Machine Learning* para grafos. As principais tarefas que direcionam as pesquisas são: Classificação de nós, classificação de grafos e predição de novas arestas.

Alguns dos principais modelos de ML que vieram a influenciar fortemente a área são aqueles que aplicaram ideias da área de Processamento de Linguagem Natural, como o SkipGram usado no word2vec [Mikolov et al. 2013], para aprender representações de nós a serem utilizados em classificadores. Exemplos desses modelos são o node2vec [Grover and Leskovec 2016] e o DeepWalk [Perozzi et al. 2014]. Alguns outros modelos que surgiram a partir desses são: struc2vec [Ribeiro et al. 2017] e o NBNE [Pimentel et al. 2019].

Uma outra ferramenta que é aplicada a grafos são as redes neurais convolucionais. Diferentemente de dados estruturados como tabelas, que possuem linhas ordenadas, e outros dados não estruturados como imagens, que possuem relações fixas entre os *pixels*, os nós de um grafo não possuem uma ordenação específica. Além disso, a quantidade de vizinhos de cada nó não é constante. Essas são as principais questões que dificultam a aplicação de redes neurais convolucionais em grafos. Entretanto, é possível adaptá-las para grafos.

A aplicação das n camadas de convoluções sobre os vizinhos de um nó forma as chamadas *Graph Neural Networks*. Esses modelos se diferenciam pela função de agregação das representações de nós vizinhos para gerar a representação de um nó e também pela função que combina as representações de todos os nós de um grafo para gerar a representação do próprio grafo [Xu et al. 2018]. Exemplos desses modelos são as *Graph Convolutional Networks* (GCN) [Kipf and Welling 2016], GraphSAGE [Hamilton et al. 2017] e *Graph Attention Networks* (GAT) [Veličković et al. 2017].

5. Expansão da Rede de Ruas de Cidades

5.1. Expansão ao Longo do Tempo

A expansão de uma cidade acontece devido a necessidade de abrigar cada vez mais pessoas que estão interessadas em morar nessa cidade por algum motivo. Quanto mais pessoas chegam, é preciso que a quantidade de residências acompanhe esse movimento e, portanto, mais espaço é ocupado, aumentando a área da cidade. Com mais espaço construído, ruas devem acompanhar a expansão para trazer mobilidade e permitir o transporte de bens, pessoas e materiais. Assim, a expansão de uma cidade como um todo está fortemente ligada ao seu processo de urbanização.

A Revolução Industrial, no início do século 19, foi um ponto muito importante para o aumento da urbanização. As taxas de urbanização nunca haviam passado de 20% dentre os principais países europeus e EUA até 1800 mas, ao final do século, o Reino Unido já chegava à taxa de 70% sendo acompanhado por Bélgica, Holanda e Alemanha. Os países da América Latina se industrializaram mais tarde, mas ultrapassaram a taxa de população urbana europeia em 1972 [Angel 2012].

A expansão de uma cidade pode sofrer interferências se ela for vista como algo a ser evitado. No século 17, a Rainha Elizabeth 1 da Inglaterra realizou inúmeras proclamações limitando a construção de novas residências nos arredores de Londres. Apesar disso, durante o período dos anos de 1800 a 2000, a população de Londres aumentou 10 vezes, de 1 milhão para 10 milhões, enquanto a sua área urbana aumentou 63 vezes, de 3.600 hectares para 230.000 hectares[Angel 2012]. Isso também mostra que a expansão da área urbana não só pode seguir mas também ultrapassar a taxa de crescimento populacional o que diminui a sua densidade.

Um exemplo recente de tentativa de interferência na expansão populacional e, conseqüentemente, das cidades, pode ser o da China. Até 2012, mais de 260 milhões de pessoas migraram de zonas rurais para os centros urbanos, constituindo, portanto, cerca de 40% da população urbana apesar de não possuírem a permissão necessária[Angel 2012].

Em contraste, a cidade de Nova Iorque, em 1811, terminava de preparar o seu plano para a ilha de Manhattan, prevendo o aumento da área construída em 7 vezes. Ao final do século, sua área realmente aumentou 7 vezes e a população 20 vezes, passando de 96 mil para 1.85 milhão. Assim, como a maior parte daquilo que já havia sido planejado fora ocupado, realizaram mais um plano ao unir os condados vizinhos do Bronx, Queens e Richmond à cidade e a prepararam para mais uma expansão de área construída de 7 vezes. Dessa forma, em 1930, sua população já chegara a quase 7 milhões. No ano 2000, a área urbana de Nova Iorque era de cerca de 6.200 Km quadrados com uma população de mais de 16 milhões[Angel 2012]. Nesse caso, a taxa de aumento da área da cidade foi menor do que o crescimento populacional.

Para acompanhar o crescimento populacional, a rede de ruas também deve crescer. Por exemplo, [Strano et al. 2012], ao analisar a expansão da rede urbana em uma área ao redor de Milão, detectou que a quantidade de nós segue uma função linear com relação a população da área estudada. Esse resultado vai ao encontro com o que [Barthelemy et al. 2013] encontrou em outro estudo analisando Paris em um período de 200 anos. Para se ter um ideia, entre 2010 e 2050, é esperado que a população urbana em países desenvolvidos cresça em 160 milhões - 100 milhões apenas nos EUA. Já para países em desenvolvimento, é esperado um crescimento de 2.6 bilhões[Angel 2012].

Dessa forma, a decisão de onde e quando estabelecer novas ruas em uma cidade é muito importante pois ela, provavelmente, definirá o traçado de transporte de pessoas pela cidade por muitos séculos[Floater et al. 2014]. Às vezes, medidas drásticas e planos audaciosos devem ser levados a frente para modernizar a rede urbana. Paris, por exemplo, durante o período Haussmann do século 19, sofreu grandes transformações. Essas transformações levaram o centro de Paris de um paradigma medieval para um mais eficiente com avenidas grandes e ruas mais seguras. Ao modelar Paris como um grafo primal, em apenas 50 anos, entre 1836 e 1888, Paris viu a sua quantidade de nós e o comprimento total de suas ruas dobrar. As transformações realizadas nessa época moldaram as próximas alterações que viriam ao estabilizar as medidas de centralidade das ruas enquanto outras novas eram construídas[Barthelemy et al. 2013].

Os casos de Nova York e Paris mostram que, quando a expansão da rede urbana está fortemente submetida a uma autoridade central que possui um claro planejamento no que tange a disposição de vias públicas, a cidade poderá usufruir de benefícios e se

prevenir de problemas futuros.

Uma das, se não a principal, características das ruas de Nova York é a sua forte disposição em grade (*grid*). Essa disposição é vista como uma das melhores para aumentar a trafegabilidade, conectividade e capacidade de caminhar para os cidadãos de uma cidade [Barrington-Leigh and Millard-Ball 2019b] [Burghardt et al. 2022]. Entretanto, o que se observou nas últimas décadas de expansão é um afastamento dessa disposição para ruas mais longas, menos interseções entre 4 vias e, portanto, cidades menos conectadas [Barrington-Leigh and Millard-Ball 2020] [Barrington-Leigh and Millard-Ball 2019b]. Entretanto, esse movimento não é uniforme ao redor do mundo. Países da América do Sul, principalmente aqueles que são ex-colônias espanholas, possuem bons níveis de conectividade já que suas cidades tendem a seguir o padrão em *grid* desde a sua colonização. Japão e China também possuem alta conectividade. Os Estados Unidos é conhecido pela baixa conectividade geral devido a influência do padrão *cul-de-sac* aplicado durante o último século. Esse padrão criou bairros cheios de ruas sem saída e fez com que as pessoas se movimentassem pela cidade priorizando o uso de automóveis privados [Barrington-Leigh and Millard-Ball 2020].

5.2. Compactas ou Espalhadas?

Atualmente, existe um debate que perdura desde a década dos anos 1990 sobre qual paradigma de desenvolvimento de cidade seria melhor: uma cidade compacta ou uma cidade mais espalhada [Ewing and Hamidi 2015].

Cidades compactas são cidades com desenvolvimento com continuidade espacial, centros fortes, usos de terra mistos, densidades média a alta, boa acessibilidade e espaços abertos permanentes. Já cidades com um certo nível de espalhamento desordenado (*sprawl*) são o contrário, ou seja, possuem desenvolvimento esparso e grande expansões de desenvolvimentos de baixa-densidade [Ewing 1997] [Ewing and Hamidi 2015] [Gordon and Richardson 1997].

A ideia da preferência de cidades compactas sobre as mais espalhadas se baseia em questões de sustentabilidade ambiental e energética além de levar em consideração o congestionamento do trânsito [Gordon and Richardson 1997]. Entretanto, esse paradigma sofre com a existência do chamado Paradoxo das Cidades Compactas. Esse paradoxo mostra que cidades compactas podem possuir correlação fraca ou limitada com sustentabilidade além de possuir baixa habitabilidade, uma crítica a sua alta densidade populacional. Assim, para uma cidade ser sustentável, seus principais ativos e a população devem estar concentrados em densidades maiores. Entretanto, para uma cidade prover uma boa experiência de vida, seus ativos e a população devem ser dispersos em densidades menores [Neuman 2005]. Apesar desse paradoxo, em um estudo mais recente na cidade de Oslo, residentes de áreas que seguem o paradigma cidade compacta parecem estar mais satisfeitos com a sua vizinhança, o que mostra que cidades compactas tem potencial para ser tanto sustentáveis quanto habitáveis [Mouratidis 2017].

Independentemente da escolha entre cidades compactas ou mais espalhadas, no que tange a conectividade e planejamento de ruas, elas parecem ter algo em comum. Neuman, ao listar as características de cidades compactas, inclui o item "alto grau de conectividade das ruas" e, para as cidades espalhadas, inclui "transporte dominado por automóveis privados" [Neuman 2005]. O caso das cidades compactas já menciona conec-

tividade mas, para as cidades espalhadas, a preferência pelo uso de automóveis privados ainda fomenta a necessidade do planejamento urbano que vise a melhora da conectividade das ruas e que facilite a locomoção dos carros para diminuir o congestionamento. Inclusive, sobre essa última questão, sobre o ponto de vista de cidades espalhadas, [Gordon and Richardson 1997] argumentam que, com a suburbanização e o desenvolvimento de vários centros, os trabalhadores não precisam todos se locomover até um único centro da cidade, o que ajudaria a desafogar o trânsito. Um outro ponto importante que os mesmos trazem na década de 1990 e que age como uma previsão do momento atual é o de que o avanço das tecnologias da informação e comunicação contribuem para que os trabalhadores não precisem ficar se locomovendo para trabalhar. Assim, para a movimentação restante, boa parte dela aconteceria entre subúrbios e não entre subúrbios e um centro.

Ainda sobre o planejamento de ruas, [Angel 2012] advoga o ressurgimento do paradigma *make-room*. Ele argumenta que, para cidades que talvez já não tenham uma taxa de crescimento populacional tão alta, pode ser interessante aplicar o paradigma compacto. Entretanto, para aquelas em pleno crescimento populacional, é necessário permitir a expansão em área das cidades para evitar inúmeros problemas. Não somente isso, também é necessário planejar e construir as ruas em áreas a serem habitadas de antemão para que se mantenha um controle da qualidade da conectividade e qualidade geral da rede e de fornecimento de serviços públicos para os futuros moradores. Sobre a densidade populacional, ele descreve que às cidades deve ser permitido aumentá-la ou diminuí-la para que ela fique em uma faixa ótima, entretanto, não fornece números para tal. Inclusive, parece não existir tal quantidade perfeita para densidade, também conhecida como *goldilocks density*, apenas um consenso geral de que a densidade ideal leva em consideração fatores específicos de cada cidade.

6. Conjunto de dados utilizado

Em 2018, [Boeing 2018b], baixou e disponibilizou o grafo primal da rede urbana de cerca de 27 mil áreas urbanas dos EUA, uma das maiores até hoje. Para tal, ele utilizou da biblioteca Python OSMnx [Boeing 2017], que ele mesmo desenvolveu, para conseguir os grafos de localidades baseado nos mapas disponíveis pelo *OpenStreetMap* [OpenStreetMap contributors 2017]. O *OpenStreetMap* é uma plataforma colaborativa onde voluntários podem demarcar ruas, construções e outros elementos no mapa do mundo e estima-se que, em 2017, ele estava mais de 80% completo [Barrington-Leigh and Millard-Ball 2019a]. Entretanto, como o objetivo do trabalho é conseguir dados representativos do mundo inteiro para servirem de treino em um modelo de *Graph Machine Learning* na próxima etapa do trabalho, é necessário que o conjunto de cidades escolhidas englobe vários países. Além disso, como [Angel 2012] explica: *“a aplicação de regras e planos que foram feitas para cidades nos EUA ou Europa não deveriam ser as mesmas para cidades de outros países no mundo menos desenvolvido, entretanto, elas ainda acontecem”*.

Já em 2021, [Boeing 2021], seguindo com o seu trabalho, conseguiu uma amostra representativa de 8914 áreas urbanas de 178 países do mundo. Ele também utilizou da biblioteca OSMnx para baixar grafos planares, direcionados e possivelmente não conectados de cidades disponíveis no *OpenStreetMap*. Ao final, o conjunto de dados possui 37 milhões de nós e 53 milhões de arestas. Além dos grafos em si, ele também computou 35

métricas utilizando também do OSMnx e de outras fontes disponíveis.

Assim, dada a representatividade dos dados e da presença de várias métricas já calculadas, esse último foi o conjunto de dados escolhido para usar como base deste trabalho. Os dados baixados foram os arquivos relativos aos grafos em si [Boeing 2020a], os indicadores [Boeing 2020b] e os metadados [Boeing 2020c]. Todos os indicadores calculados e suas descrições podem ser vistos na Seção 7.

7. Métricas encontradas e computadas sobre os grafos das cidades

Os paradigmas citados e o debate de qual aplicar nas cidades mostra que fatores como densidade, urbanização e planejamento prévio da rede urbana são essenciais para a qualidade geral da vida da população. Além disso, a modelagem da rede de ruas como um grafo permite a criação e uso de outras métricas que são aplicadas em redes.

Dessa forma, existem muitas métricas que podem caracterizar o espaço urbano. [Boeing 2018a] cria uma tipologia dessas métricas e as divide em 5 grupos diferentes: Temporal, Visual, Espacial, Escala e Conectividade. Dado que o objetivo do trabalho final é prever novas ruas com base no grafo das cidades, resolvi focar em encontrar e aplicar medidas de conectividade, apesar de outras mais gerais também terem sido computadas.

Ao total, foram encontradas 46 métricas diferentes criadas e aplicadas sobre o grafo primal das redes urbanas. Das 46, 42 podem ser vistas, junto com sua descrição e referências, na Tabela 2. Dessas 42, 35 já haviam sido computadas por [Boeing 2021] no conjunto de dados descrito e escolhido na Seção 6. O objetivo aqui, portanto, é o de aumentar, o tanto quanto possível, a quantidade de indicadores disponíveis para que possam ser utilizados para filtrar as 8914 cidades do conjunto de dados já mencionado àquelas consideradas planejadas.

Table 2: Métricas encontradas com descrição e referências

Métrica	Descrição	Referências
Efeito de <i>clustering</i> médio	Mede quanto os vizinhos de um nó formam um clique, ou seja, quão conectados estão os vizinhos de um nó.	[Boeing 2021]
Circuito (<i>circuity</i>)	Proporção do comprimento de rua para comprimento de ruas retas	[Boeing 2021]
Elevação de nós	Elevação de nós com relação ao mar (metros)	[Boeing 2021]
Inclinação da rua	Inclinação das ruas	[Boeing 2021]
<i>Bearing</i> da rua	<i>Bearing</i> da rua	[Boeing 2021]
Quantidade de interseções	Quantidade de cruzamentos totais no grafo. Dessa forma, exclui nós que representam ruas sem saída	[Boeing 2021]
Grau médio dos nós	Soma dos graus de todo nó no grafo dividido pela sua quantidade. Só leva em consideração arestas não direcionadas	[Boeing 2021]

Table 2: Métricas encontradas com descrição e referências

Métrica	Descrição	Referências
Comprimento das arestas	Comprimento associado a cada uma das arestas	[Boeing 2021]
Número de nós	Quantidade total de nós no grafo. Os nós representam interseções e ruas sem saída.	[Boeing 2021]
<i>Pagerank</i> máximo de um nó	Máximo valor de <i>PageRank</i> dentre todos os nós no grafo	[Boeing 2021] [Page et al. 1999]
Proporção de interseções de 4 vias	A porcentagem das interseções que ligam 4 arestas	[Boeing 2021]
Proporção de interseções de 3 vias	A porcentagem das interseções que ligam 3 arestas	[Boeing 2021]
Proporção de nós representando ruas sem saída	A porcentagem de nós que representam ruas sem saída	[Boeing 2021]
Proporção de arestas que são <i>self loops</i>	Proporção de arestas que retornam para o próprio nó de origem	[Boeing 2021]
Linearidade	Calculado como $1 - \textit{circuitry}$	[Boeing 2021]
Quantidade de segmentos de ruas	Quantidade de arestas no grafo não direcionado	[Boeing 2021]
Região do mundo	Região do mundo da cidade	[Boeing 2021]
Sub Região do mundo	Sub Região do mundo da cidade	[Boeing 2021]
País	País da cidade	[Boeing 2021]
População	Valor estimado em 2015	[Boeing 2021]
Área construída total	Área em km^2 que contém construções na cidade. Valor estimado em 2015	[Boeing 2021] [Angel 2012]
Área total	Área total em km^2 da cidade	[Boeing 2021]
Densidade Populacional	Densidade populacional com relação à área total construída	[Angel 2012] [Angel et al. 2021]
Densidade de interseção	Número de interseções por km^2 na cidade	[UNH 2016] [Boeing 2018b]
Proporção orgânica	Proporção de interseções entre 3 vias e de ruas sem saída no grafo.	[Courtat et al. 2011] [Rui et al. 2013]
Coeficiente de Malha (<i>Meshedness Coefficient</i>)	Mede quanto um grafo se parece com uma árvore	[Courtat et al. 2011]

Table 2: Métricas encontradas com descrição e referências

Métrica	Descrição	Referências
Entropia de Shannon	Aplicado sobre os tipos hierárquicos de ruas da cidade.	[Xie and Levinson 2007]
<i>ringness</i> (ϕ_{ring})	Comprimento de vias arteriais em subgrafos considerados anéis sobre o comprimento total de vias arteriais no grafo	[Xie and Levinson 2007]
<i>webness</i> (ϕ_{web})	Comprimento de vias arteriais em subgrafos considerados teias sobre o comprimento total de vias arteriais no grafo	[Xie and Levinson 2007]
<i>circuitness</i> ($\phi_{circuit}$)	Soma do <i>ringness</i> com <i>webness</i>	[Xie and Levinson 2007]
<i>treeness</i> (ϕ_{tree})	$1 - circuitness$	[Xie and Levinson 2007]
Densidade de nós	Quantidade de nós por km^2 . Nós envolvem tanto ruas sem saída quanto interseções	[Boeing 2018b]
Densidade de arestas	Divide o comprimento total das arestas direcionadas pela área (km^2)	[Boeing 2018b]
Densidade de ruas	Divide o comprimento total das arestas direcionadas e não direcionadas pela área (km^2)	[Boeing 2018b]
Ruas médias por nó	Parecido com o grau mas também leva em consideração nós que representam ruas sem saída	[Boeing 2018b]
Conectividade do grafo	A quantidade mínima de nós que, se removidos, desconectam o grafo	[Boeing 2018b]
Conectividade média dos nós	A quantidade média de nós que se deve remover para desconectar dois nós aleatórios	[Boeing 2018b]
<i>Betweenness Centrality</i>	Medida de quantos dos menores caminhos do grafo passam por cada nó.	[Porta et al. 2005] [Crucitti et al. 2006] [Cao et al. 2016] [Boeing 2018b]
Distância característica	A distância média dos menores caminhos calculada sobre todos os pares de nós na rede	[Porta et al. 2005]
Eficiência Global	Variante da Distância Característica que funciona sobre grafos não conectados	[Latora and Marchiori 2001] [Porta et al. 2005]
Nós por km de rua	Quantidade total de nós dividido pelo comprimento total das ruas	[Burghardt et al. 2022]
SNDi	PCA aplicado sobre algumas outras métricas para estimar o <i>sprawl</i>	[Barrington-Leigh and Millard-Ball 2019b]

Para 5 das métricas da Tabela 2, variantes também foram computadas. Essas variações podem ser vistas na Tabela 3.

Table 3: Variantes computadas para algumas métricas

Métrica	Variantes
Efeito de <i>clustering</i> médio	Arestas sem peso e direcionadas; Arestas sem peso e não direcionadas; Arestas com peso e direcionadas; Arestas com peso e não direcionadas;
Elevação	Variação interquartil (IQR); Média; Mediana; Faixa; Desvio Padrão
Inclinação	Média; Mediana
Interseção	Quantidade total; Quantidade após agregar nós em distância menor do que 10 metros geometricamente; Quantidade após agregar nós em distância menor do que 10 metros topologicamente;
Comprimento	Média; Mediana; Total
<i>Bearing</i>	Entropia; Ordem
<i>Betweenness Centrality</i>	Média; Máximo

Além disso, outras três métricas foram encontradas disponíveis nas bibliotecas Networkx [Hagberg et al. 2008] e OSMnx [Boeing 2017] do Python:

- Conectividade de arestas: Retorna o conjunto de arestas de menor cardinalidade que desconectam o grafo
- Quantidade de interseções entre n vias: Essa variável apresentou resultados para variações de $n \in N$ da forma $n \in [1, 28]$.
- Proporção de interseções entre n vias: Essa variável apresentou resultados para variações de $n \in N$ da forma $n \in [1, 28]$ menos para os valores $\{1, 3, 4\}$.

Uma última foi computada a partir da relação entre a área total construída e a área total da cidade, chamada de "proporção de área construída".

Dessa forma, contando todas as métricas encontradas e suas variações, temos 111 métricas encontradas. Entretanto, as seguintes 12 métricas não puderam ser computadas:

1. Entropia de Shannon: Era necessário que todas as arestas estivessem associadas a um tipo de via, por exemplo, arterial ou coletora;
2. *ringness*: Também era necessário que todas as arestas estivessem associadas a um tipo de via;
3. *webness*: Também era necessário que todas as arestas estivessem associadas a um tipo de via;
4. *circuitness*: Precisa das métricas *ringness* e *webness*;
5. *treeness*: Precisa da métrica *circuitness*;
6. Conectividade do grafo: Como existem ruas sem saída, ao desconectar o nó relativo à entrada dessa rua, o grafo já passaria a ser desconectado. Assim, a resposta sempre era igual a 1;
7. Conectividade de arestas: Sofreu com o mesmo problema da métrica Conectividade do grafo. Ao retirar a aresta que liga o nó da rua sem saída e o nó do início

da rua, o grafo já se tornava desconectado. Portanto, a resposta também sempre era igual a 1;

8. Conectividade média dos nós: Essa métrica leva em consideração os menores caminhos entre todos os pares de nós possíveis, ela se tornou muito cara para computar;
9. Distância característica: Essa métrica também leva em consideração os menores caminhos entre todos os pares de nós possíveis, portanto, era muito cara para computar;
10. Eficiência global: Essa métrica também leva em consideração os menores caminhos entre todos os pares de nós possíveis, portanto, era muito cara para computar;
11. *Betweenness Centrality*: Essa métrica também leva em consideração os menores caminhos entre todos os pares de nós possíveis, portanto, era muito cara para computar. Como essa métrica possui duas variações indicadas na Tabela 3, ela conta como duas métricas não computadas.
12. SNDi: Era necessário calcular as métricas, ainda não citadas, de Dendricidade e Sinuosidade. Como não foram encontradas funções disponíveis para calcular tais métricas, o SNDi não pôde ser calculado.

Todas essas métricas foram calculadas sobre 8909 das cidades do conjunto de dados escolhido. As 5 maiores cidades com relação ao número de nós foram deixadas de lado devido a restrições de memória. Essas cidades deixadas de fora foram: *tokyo*, *jakarta*, *osaka_kyoto*, *nagoya* e *mexico_city*.

Assim, a computação de todas essas métricas e variações constituem um *dataset* [de Campos 2022b] com 98 *features*, além de outras 4 informações de identificação dos dados como nome da cidade e "id", sobre cada cidade do conjunto de dados escolhido.

8. Escolha de cidades planejadas

8.1. Explorando os dados

Com tantas *features* disponíveis, para se ter uma ideia rápida da relação entre elas, resolvi visualizar a matriz de correlação entre 38 das *features*. A correlação está mostrada na Figura 1. Nela, é possível perceber a existência de duas outras *features* não mencionadas, mas que foram criadas temporariamente: *big*, classificando as cidades como 1 se elas possuírem mais de 20 mil nós e 0 caso contrário; e *orgânica*, classificando as cidades como *True* se a sua proporção orgânica for menor do que 0.5 e *False* caso contrário.

Nessa figura, é possível observar algumas correlações já esperadas:

- **Coefficientes de *clustering***: As variáveis de coeficiente de *clustering* (as que começam com *cc*), que indicam quanto os vizinhos de um nó/interseção estão conectados entre si, possuem alta correlação umas com as outras. Entretanto, elas não possuem correlação com variáveis de densidade de interseção ou nós. Isso pode indicar que, apesar de existirem cidades mais densas, com interseções mais próximas umas das outras, essas interseções não estão, necessariamente, conectadas umas às outras.
- **Circuito**: A variável que indica quão parecido com uma linha reta é o caminho entre dois nós (*circuitry*) possui alta correlação negativa com a sua inversa *straightness*.

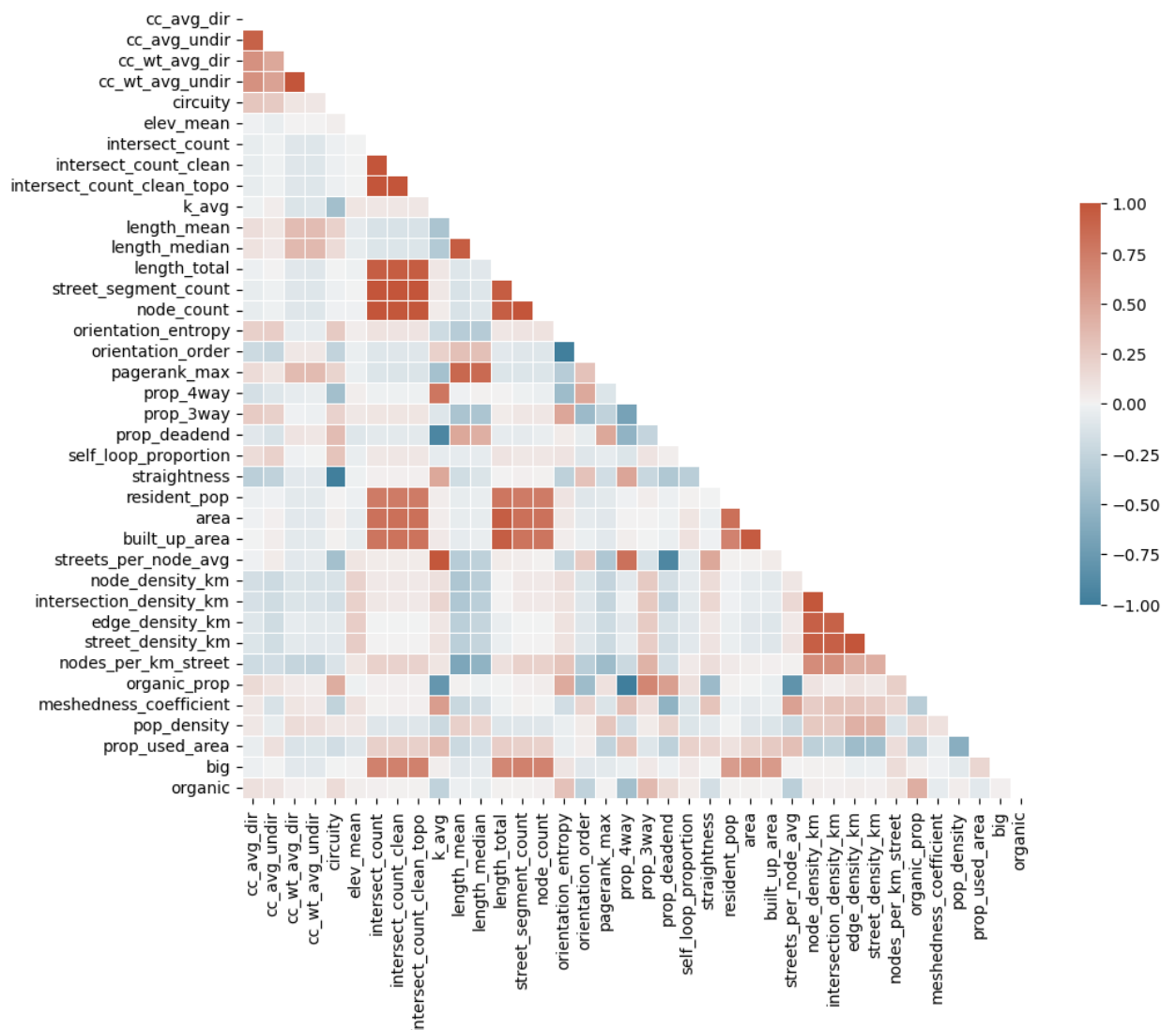


Figure 1. Correlação entre 38 das *features* computadas

- Contagem de interseções:** As variáveis de contagem de interseções (as que começam com `intersect_count`) possuem alta correlação umas com as outras. Além disso, também possuem alta correlação com: contagem de nós (`node_count`), já que nós englobam interseções e representações de ruas sem saída; contagem de segmentos de ruas (`street_segment_count`), já que, para cada interseção, existem pelo menos duas arestas representando dois segmentos de ruas; comprimento total das ruas (`length_total`), já que, com mais segmentos de ruas, tem-se maior comprimento total da malha de ruas; população residente (`resident_pop`), já que mais interseções representam mais ruas e, consequentemente, cidades maiores que abrigam mais pessoas; área (`area`) e área total construída (`built_up_area`), já que, com mais área construída, o que precisa de mais área total para a cidade, é possível ter mais construções que necessitam de ruas aumentando, consequentemente, a quantidade de interseções; cidade grande (`big`), já que foram consideradas cidades grandes aquelas com mais de 20 mil nós;

- **Grau médio:** A variável indicando o grau médio dos nós (*k_avg*) possui alta correlação positiva com: quantidade de ruas média por nó (*streets_per_node_avg*), já que, quanto maior a média de ruas por nó, maior a média do grau; proporção de interseções de 4 vias (*prop_4way*), já que, quanto maior essa proporção, mais nós se conectam a 4 outros nós, o que aumenta o grau médio. Além disso, possui alta correlação negativa com: proporção orgânica (*organic_prop*), já que, quanto maior esse valor, mais os nós tendem a ter conexões com 1 ou 3 outros nós, o que diminui o grau médio; proporção de ruas sem saída (*prop_deadend*), já que, quanto mais ruas sem saída, significa que os nós se conectam a apenas um outro nó, diminuindo o grau médio;
- **Comprimento total:** A variável indicando o comprimento total das ruas da rede (*length_total*), possui alta correlação com: área (*area*) e área total construída (*build_up_area*), já que, quanto maior a área disponível e construída, mais ruas são necessárias; população residente (*resident_pop*), já que, quanto mais pessoas, mais ruas são necessárias para as construções que devem abrigá-las;
- **Contagem de segmentos de ruas e quantidade de nós:** A variável indicando a quantidade de segmentos de ruas ou arestas (*street_segment_count*) e a que indica a quantidade total de nós (*node_count*) possuem as mesmas correlações que a variável *length_total* e por motivos semelhantes.
- **Proporção de interseções de 4 vias:** Essa variável (*prop_4way*) possui alta correlação relativa com proporção orgânica (*organic_prop*), já que, quanto mais interseções entre 4 vias, menos existem entre 3 vias e ruas sem saída, o que diminui a proporção orgânica. Além disso, possui alta correlação positiva com a quantidade de ruas média por nó (*streets_per_node_avg*), já que, sendo uma variável *proxy* para o grau médio dos nós, quanto maior for o seu valor, maior será o *prop_4way*.
- **Proporção de ruas sem saída:** Essa variável (*prop_deadend*) possui alta correlação negativa com quantidade de segmentos de ruas médio por nó (*street_per_node_avg*) já que, quanto mais ruas sem saída, menor é o grau dos nós, na qual *street_per_node_avg* é uma *proxy*.
- **Variáveis de densidade:** Essas variáveis (terminadas em *density_km*) possuem alta correlação positiva entre elas mesmas já que, quanto mais ruas em uma área, maior a quantidade de interseções também.
- **Proporção da área usada:** Essa variável mede quanto da área total disponível para a cidade já foi usada. Ela possui uma correlação negativa com a densidade populacional, indicando que quanto mais área é utilizada, menor a densidade populacional e vice-versa.

Além disso, existem algumas correlações que podem ser consideradas inesperadas:

- **Proporção orgânica:** Essa variável (*organic_prop*) não possui correlação com a densidade de nós/interseções (*node_density_km* e *intersection_density_km*), o que pode indicar que tanto cidades mais densas quanto mais espalhadas nesses sentidos podem ser mais orgânicas ou não. Esses movimentos podem ser vistos na Figura 2. Essa variável também não apresenta correlação com a quantidade de nós (*node_count*), o que indica que não existe uma regra com relação ao tamanho de uma cidade e o seu planejamento retratado pela proporção orgânica. O mesmo vale com as variáveis de área e população residente.

- **Coefficiente de *clustering*:** Essa variável (as que começam com *cc*) indica quanto os vizinhos de um nó/interseção estão conectados entre si. Entretanto, ela não possui correlação com as variáveis *prop_deadend*, *prop_3way* e *prop_4way*. Isso indica que, apesar de existirem cidades com diferentes proporções desses tipos de interseções, o coeficiente de *clustering* não segue uma regra clara.
- **Máximo valor de *pagerank*:** Essa variável possui alta correlação positiva com variáveis indicando comprimento médio e total das ruas. É possível ver essa correlação na Figura 3. Não está claro o porquê desta situação.
- **Quantidade de interseções:** Essa variável não possui correlação com variáveis de densidade de interseção, nós ou arestas. Isso indica que a densidade tem pouca influência na quantidade de interseções, sendo mais afetado pela área total da cidade.
- **Densidade populacional:** Essa variável não possui correlação com a quantidade de interseções, comprimento total das ruas da rede e nenhuma outra variável.

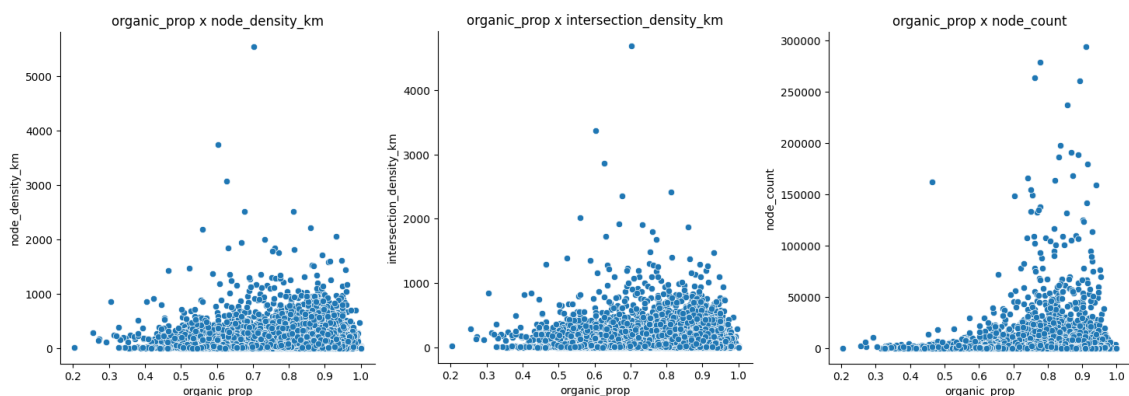


Figure 2. Correlações inesperadas envolvendo a métrica proporção orgânica. No caso, a questão é que não existe correlação aparente.

8.2. Proporção Orgânica

De acordo com o que foi dito na Seção 5 sobre expansão de cidades, parece que os principais eixos de análise são: Densidade Populacional das cidades e Conectividade da rede urbana.

Uma variável que atual no eixo da Conectividade é a proporção orgânica. A proporção orgânica mede a proporção de interseções que ligam 3 ruas ou 1 (ruas sem saída). Quanto mais próxima de 1, teoricamente, mais desconectada a cidade será. Quanto mais próxima de 0, significa que, provavelmente, essa cidade possui uma alta proporção de interseções entre 4 vias, que, como já foi dito, é benéfica no geral para as cidades. Dessa forma, essa métrica atua como uma indicadora de conectividade do grafo e já abrange as variáveis de proporção de interseção entre n vias.

A distribuição dos valores da proporção orgânica no *dataset* pode ser vista na Figura 4. É possível ver que a maioria das cidades possui proporção orgânica maior do que 0.6. Se definirmos que uma cidade não é orgânica caso o seu valor para essa métrica seja menor do que 0.5, 184 cidades caem nessa definição. Também é interessante notar que existem 33 cidades totalmente orgânicas, a maioria delas na Índia. A distribuição de

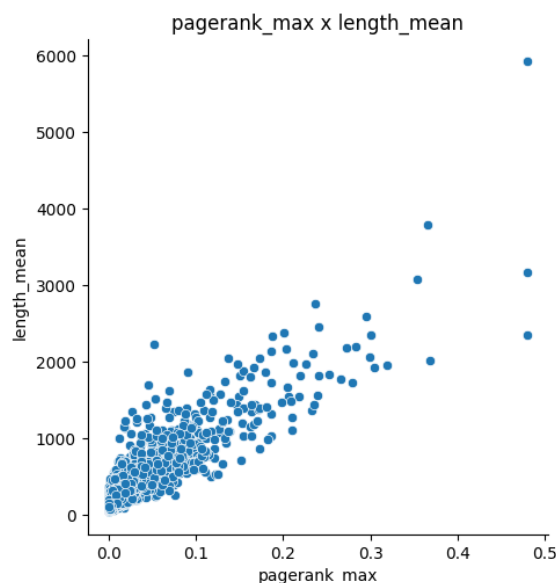


Figure 3. Correlação inesperada entre PageRank máximo e o comprimento médio total das ruas.

idades não orgânicas por região do mundo pode ser vista na Figura 5. Assim, chama a atenção a não existência de cidades não orgânicas da Europa, Oceania e América do Norte.

Na Figura 6 é possível ver a distribuição do valor de proporção orgânica para cada região do mundo. Pode-se perceber que as cidades da América Latina e Caribe possuem menor média de proporção orgânica. Isso vai ao encontro com o que foi dito por [Barrington-Leigh and Millard-Ball 2020] que disse que Países da América do sul possuíam bons níveis de conectividade. China e Japão, apesar de dito que também possuem alta conectividade, estão no grupo da Ásia que possui uma das maiores médias dessa métrica. Chama a atenção a alta variabilidade para a América Latina, África e Ásia, enquanto Oceania possui a mais baixa variabilidade.

8.3. Densidade de Interseções

A densidade de interseções é outra métrica que atua no eixo da Conectividade. Essa métrica é uma boa indicadora de quão provável é uma pessoa poder caminhar ou pedalar até outro lugar, além de contribuir para a existência de mais rotas entre dois pontos em uma cidade, o que diminui o congestionamento. Um *Report* da ONU mostra empiricamente que um valor ao redor de 100 interseções por km^2 marca um nível suficiente de conectividade das ruas [UNH 2016].

A Figura 7 mostra a distribuição desta métrica para todas as cidades. É possível perceber que a maior parte das cidades possuem valor de interseção menor do que 200. Entretanto, aquelas que possuem valor ao redor de 100 não são tantas assim. Além disso, dentre as cidades definidas como orgânicas, apenas cinco estão entre o valor de 80 e 120 de densidade de interseção.

A Figura 8 mostra que não existe relação aparente entre a proporção orgânica e a densidade de interseção.

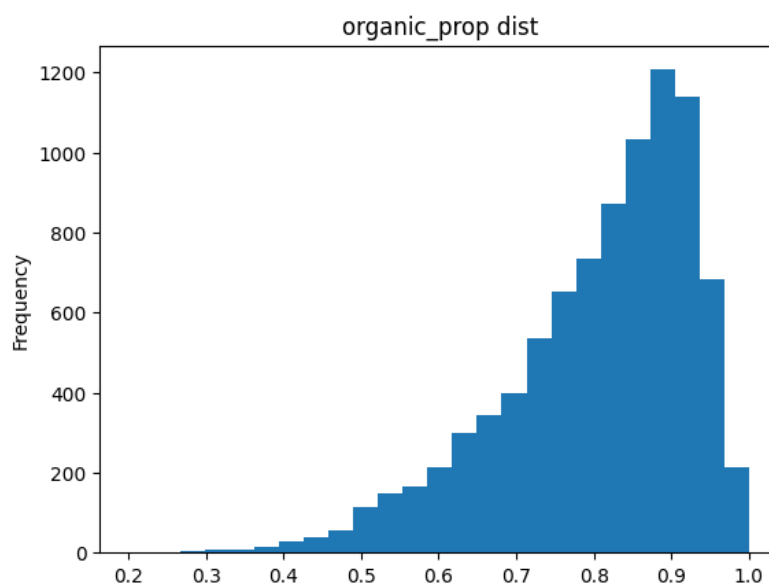


Figure 4. Distribuição da *feature* Proporção Orgânica

8.4. Alta proporção de interseção entre 3 vias

De acordo com [Barrington-Leigh and Millard-Ball 2019b], já foi mostrado que uma rede em *grid* não é a única que consegue atingir alta conectividade. Existem lugares com junções de 3 ruas que possuem essa característica.

A Figura 9 mostra a distribuição da proporção de interseções entre 3 vias. Essa figura mostra que as cidades parecem estar centralizadas ao redor do valor de 0.65. Entretanto, ainda existem cidades com valores mais extremos como maior do que 0.85. Esse tipo de cidade ainda pode estar disposta em um layout de forma que a sua conectividade permaneça alta. Naturalmente, essas cidades terão um valor de proporção orgânica alta, já que a proporção de interseções entre 3 vias fica no numerador do cálculo da proporção orgânica. Entretanto, ao analisar a distribuição da densidade de interseção dessas cidades, parece que elas estão centradas ao redor de 160.

8.5. Definindo tipos de cidades

Para finalmente definir quais cidades serão escolhidas como sendo planejadas, serão usadas as métricas: proporção orgânica, circuito, coeficiente de malha e densidade de interseção. Densidade populacional ficou de fora pois, como já dito anteriormente, não existe um consenso sobre uma faixa recomendada para essa métrica, apenas abstrações do tipo "o bastante". Assim, fica difícil escolher um valor para essa métrica.

O *Meshedness Coefficient* é outra métrica que trata de conectividade. Quanto mais próxima de 1, mais o grafo é conectado e, quanto menor for, mais o grafo se parecerá com uma árvore. Essa métrica e a proporção orgânica foram usadas em conjunto para modelar e descrever a estrutura geral da rede de ruas de cidades [Courtat et al. 2011]. Dessa forma, faz sentido continuar usando-as em conjunto para ajudar a caracterizar cidades como planejadas.

O circuito é uma métrica de eficiência da rede. Ele mede quão aproximado, em média, a distância no grafo entre dois nós quaisquer se assemelha com uma linha reta.

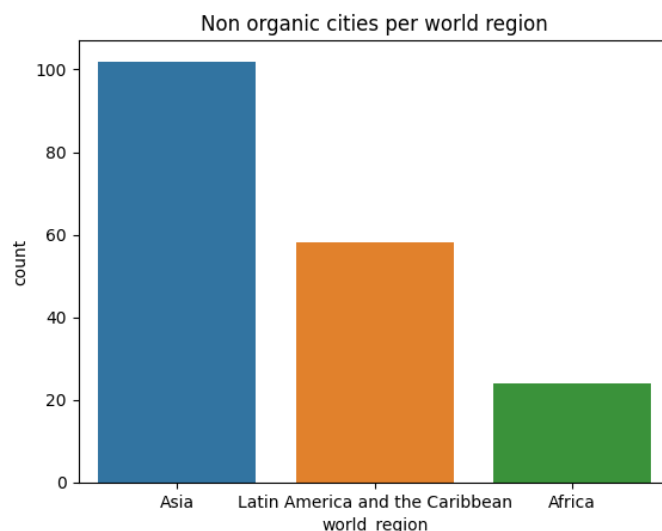


Figure 5. Quantidade de cidades definidas como não orgânicas (*organic.prop* < 0.5) por região do mundo

Isso é possível de calcular pois todos os nós tem posições fixas no mundo real, ou seja, o grafo carrega elementos geográficos como a posição do nó no mundo. Quanto menor o valor para essa métrica, ou seja, mais próximo de 1, mais eficiente é a rede.

O primeiro tipo de cidade definido são aquelas com baixa proporção orgânica e coeficiente de malha alta. Mais especificamente, a proporção orgânica deve ser menor do que 0.6 e o coeficiente de malha deve ser maior ou igual a 0.9. Esse tipo de cidade tenta capturar aquelas cuja rede de ruas é fortemente disposta em *grid* e é altamente conectada. No total, 265 caíram nessa definição.

O segundo tipo de cidade definido são aquelas que não fazem parte do tipo um, possuem uma faixa de aceitação maior de proporção orgânica (maior que 0.65) e coeficiente de malha (maior ou igual a 0.75) mas devem possuir densidade de interseção na faixa de [80, 120]. Essas cidades podem ter menor presença da disposição em *grid*, entretanto, a densidade de interseção em um nível próximo ao recomendado tem o objetivo de aproximar essas cidades de uma boa conectividade. No total, 17 cidades caíram nesse grupo.

O terceiro e último grupo tem o objetivo de abordar a questão de que cidades com alta proporção de interseções entre 3 vias ainda assim podem ter alta conectividade. Dessa forma, foram consideradas as cidades que não são do tipo 1 nem do tipo 2, possuem proporção de interseção de 3 vias maior que 0.85 e valor de circuito menor do que 1.05. O fator circuito está presente para tentar manter o nível de eficiência das cidades próximo do ótimo. No total, 31 cidades fizeram parte desse grupo.

A Figura 10 mostra a quantidade de cidades escolhidas por região do mundo. Nela, consegue-se ver a predominância de cidades da América Latina, África e Ásia com poucas da Europa e América do Norte e nenhuma da Oceania. A falta de cidades da Oceania não é surpresa pois a Figura 6 já mostrou que essa região é a que possui maior média, além de concentrada, de proporção orgânica. Entretanto, mesmo com a chance dada pelo tipo de cidades número 3, as cidades dessa região não conseguiram entrar, o

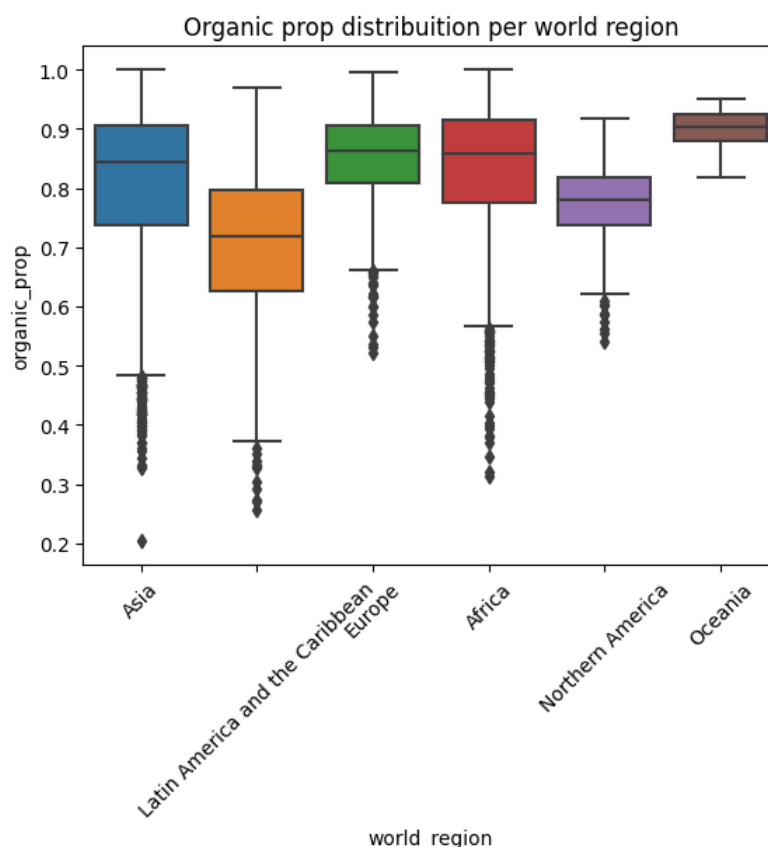


Figure 6. Distribuição do valor da proporção orgânica por região do mundo

que mostra que, além de altamente orgânicas, elas têm valor de circuito não próximo do ótimo.

Talvez a surpresa na Figura 10 seja a quantidade de cidades do continente Africano. Na Figura 6, a África é uma das regiões com maior média de proporção orgânica. Entretanto, das 84 cidades africanas escolhidas, 73 são do tipo 1 que correspondem exatamente àquelas únicas cidades africanas que possuem proporção orgânica abaixo de 0.6.

No caso da Ásia, das 75 cidades escolhidas, 49 são do tipo 1 e 21 do tipo 3. Assim, essa região é a que mais possui representantes do tipo 3. Já a América Latina teve 140 cidades escolhidas sendo 133 delas do tipo 1, a região com mais representantes desse tipo. Isso mais uma vez vai ao encontro do que já foi dito sobre as cidades da América Latina permanecerem com a forte influência colonial de desenvolvimento em *grid*.

A Figura 11 mostra as cidades que foram escolhidas sob uma visão da relação entre a proporção orgânica e o coeficiente de malha. É possível perceber claramente o recorte realizado pelas cidades do tipo 1 como aquelas com baixa proporção orgânica e alto coeficiente de malha. Além disso, é possível identificar que as cidades do tipo 2 ficam ao redor das cidades do tipo 1, de acordo com a sua definição. Já as cidades do tipo 3 são as que fogem do padrão e se distanciam das outras duas classes devido ao seu alto valor de proporção orgânica. Apesar de existirem muitas cidades não escolhidas englobando as cidades do tipo 3, a restrição do valor de circuito conseguiu filtrar em muito as cidades.

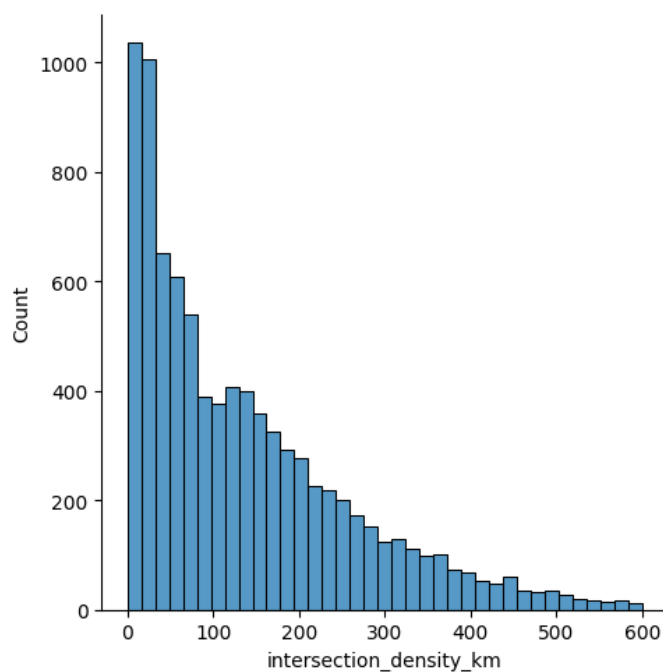


Figure 7. Distribuição do valor da densidade de interseção em km^2 para todas as cidades

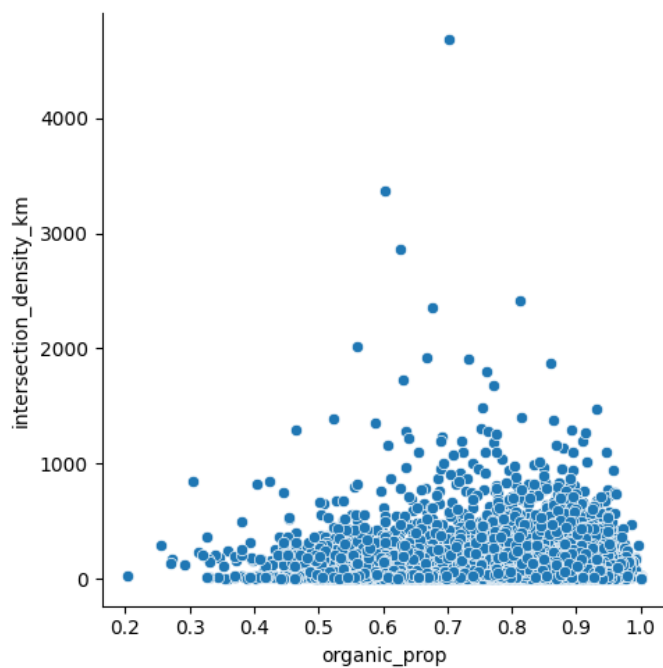


Figure 8. Relação entre proporção orgânica e densidade de interseção

O arquivo contendo o conjunto final de cidades consideradas planejadas junto com as suas *features* está disponível no GitHub [de Campos 2022a].

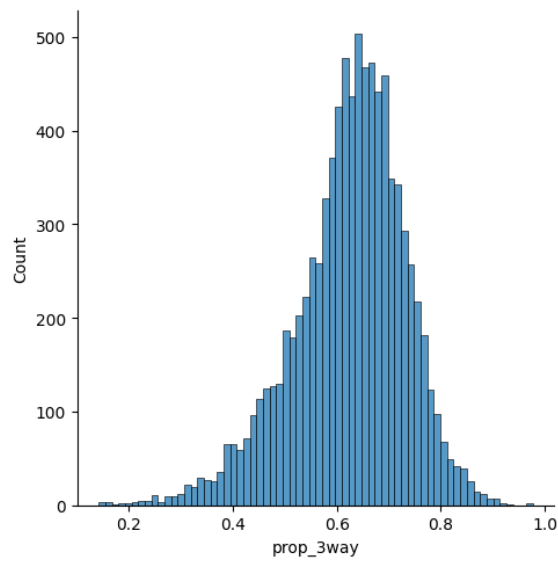


Figure 9. Distribuição do valor da métrica "proporção de interseções entre 3 vias"

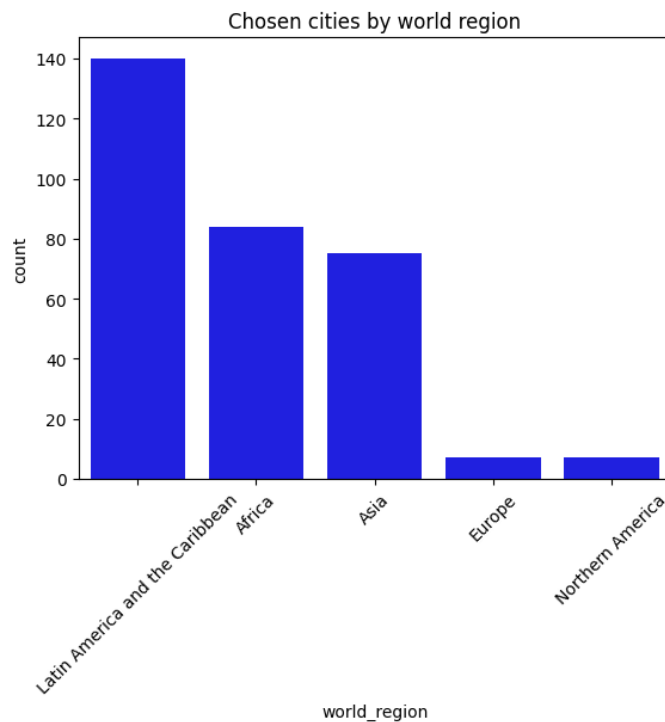


Figure 10. A quantidade de cidades escolhidas como planejadas por região do mundo

9. Considerações Finais e Próximos passos

Cidades se expandem fortemente baseado no seu processo de urbanização. Assim, quando a sua população cresce, a sua rede de ruas também cresce para permitir o transporte de cargas e pessoas. É estimado que o mundo já atingiu 8 bilhões de habitantes, sendo que a maior parte desse crescimento se deu em países em desenvolvimento. Dessa forma, as cidades deverão se preparar para acomodar todas as pessoas que ainda estão por vir nesse

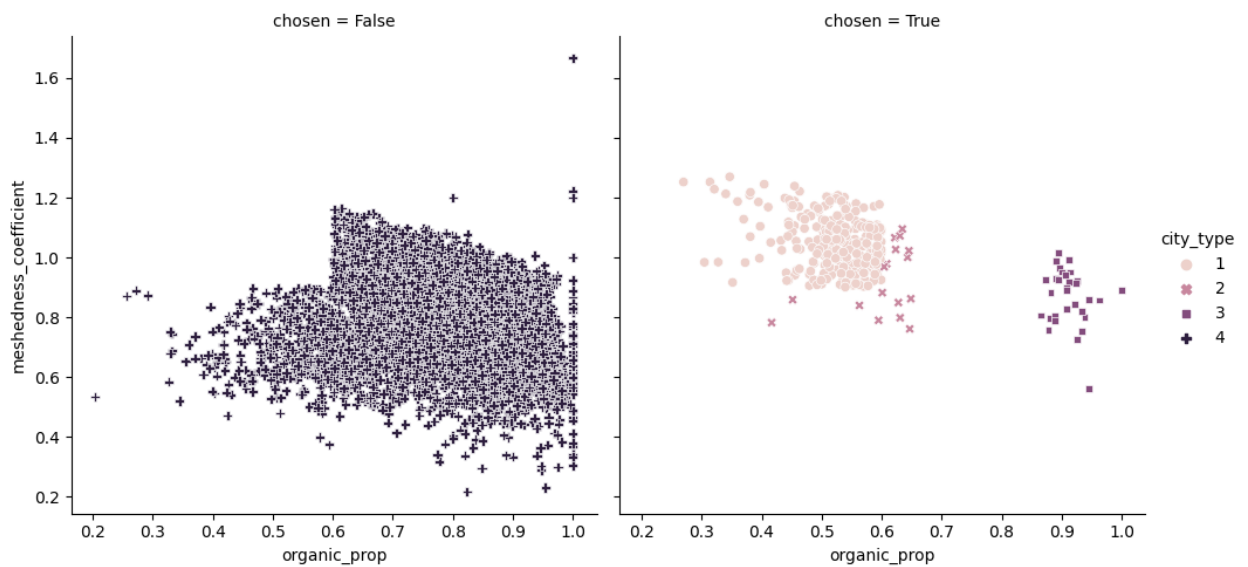


Figure 11. A relação entre proporção orgânica e o coeficiente de malha para as cidades que foram escolhidas (tipos 1, 2 e 3) e aquelas que não foram (tipo 4).

movimento de expansão.

A forma na qual esse preparo acontecerá ainda é um debate. Cidades compactas podem ser mais sustentáveis, entretanto sofrem com críticas relacionadas à sua habitabilidade. Nesse meio, existem aqueles que preferem que cidades em pleno crescimento se permitam aumentar a sua área enquanto mantêm as densidades populacionais em níveis razoáveis de acordo com a necessidade. Entretanto, o que parece ser um consenso é que cidades que planejam a sua rede de ruas de antemão evitam problemas futuros ou grandes intervenções corretivas.

Uma cidade planejada parece estar relacionada com a conectividade e eficiência de suas ruas. Assim, de acordo com a disponibilidade de métricas que sirvam de comparação entre as cidades, é possível sugerir tipos de cidades que pareçam ser caracterizadas como planejadas.

Este trabalho aumentou a quantidade de *features* já aplicadas a um *dataset* existente representativo do globo com o objetivo de filtrar essas cidades. As cidades escolhidas serão utilizadas, em um trabalho futuro, como entradas para um modelo de *Graph Machine Learning* para prever e sugerir novas ruas em cidades não consideradas planejadas com base nas que foram consideradas planejadas. A proporção orgânica, coeficiente de malha, circuito e densidade de interseção foram os atributos escolhidos que serviram de base para definir três tipos diferentes de cidades planejadas. Com esses tipos definidos, foram filtradas as cidades que se adequassem em algum desses tipos do *dataset* base.

Outros trabalhos podem tentar computar as métricas que deixaram de ser executadas por restrições computacionais para aumentar ainda mais o conjunto de *features* associados ao *dataset* de cidades. Além disso, podem tentar definir suas próprias regras de cidades planejadas baseadas nessas *features* e outras pesquisas na literatura.

Todo o código e dados estão disponíveis publicamente no GitHub.

References

- (2016). The fundamentals of urbanization evidence base for policy making. White paper, United Nations Human Settlements Programme (UN-Habitat), P. O. Box 30030, 00100 Nairobi GPO KENYA.
- Agryzkov, T., Oliver, J. L., Tortosa, L., and Vicent, J. F. (2017). Different typer of graphs to model a city. In *Computational Methods and Experimental Measurements XVIII*. WIT Press.
- Angel, S. (2012). *Planet of Cities*. Lincoln Institute of Land Policy, Cambridge, MA.
- Angel, S., Lamson-Hall, P., and Blanco, Z. G. (2021). Anatomy of density: measurable factors that constitute urban density. *Buildings and Cities*, 2(1):264–282.
- Barrington-Leigh, C. and Millard-Ball, A. (2019a). Correction: The world’s user-generated road map is more than 80% complete. *PLOS ONE*, 14(10):e0224742.
- Barrington-Leigh, C. and Millard-Ball, A. (2019b). A global assessment of street-network sprawl. *PLOS ONE*, 14(11):e0223078.
- Barrington-Leigh, C. and Millard-Ball, A. (2020). Global trends toward urban street-network sprawl. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES*, 117(4).
- Barthelemy, M., Bordin, P., Berestycki, H., and Gribaudo, M. (2013). Self-organization versus top-down planning in the evolution of a city. *Scientific Reports*, 3(1).
- Boeing, G. (2017). OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, 65:126–139.
- Boeing, G. (2018a). Measuring the complexity of urban form and design. *URBAN DESIGN International*, 23(4):281–292.
- Boeing, G. (2018b). A multi-scale analysis of 27, 000 urban street networks: Every US city, town, urbanized area, and zillow neighborhood. *Environment and Planning B: Urban Analytics and City Science*, 47(4):590–608.
- Boeing, G. (2020a). Global Urban Street Networks GraphML.
- Boeing, G. (2020b). Global Urban Street Networks Indicators.
- Boeing, G. (2020c). Global Urban Street Networks Metadata.
- Boeing, G. (2021). Street network models and indicators for every urban area in the world. *Geographical Analysis*, 54(3):519–535.
- Burghardt, K., Uhl, J. H., Lerman, K., and Leyk, S. (2022). Road network evolution in the urban and rural united states since 1900. *Computers, Environment and Urban Systems*, 95.
- Cao, H., Liu, H., Zhao, F., Li, Y., and Du, M. (2016). The evaluation of node importance in urban road network based on complex network theory. *MATEC Web of Conferences*, 61:04026.
- Courtat, T., Gloaguen, C., and Douady, S. (2011). Mathematics and morphogenesis of cities: A geometrical approach. *Phys. Rev. E*, 83:036106.

- Crucitti, P., Latora, V., and Porta, S. (2006). Centrality in networks of urban streets. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 16(1):015113.
- de Campos, D. S. (2022a). Chosen cities with 101 *features* computed. https://github.com/Pendulun/POC1/blob/main/code/analiseIndicadoresMundiais/data/chosen_cities.csv. Online available. Last update: 12/13/2022.
- de Campos, D. S. (2022b). Final data with 101 *features* computed. https://github.com/Pendulun/POC1/blob/main/code/analiseIndicadoresMundiais/data/final_df.csv. Online available. Last update: 12/13/2022.
- Ewing, R. (1997). Is los angeles-style sprawl desirable? *Journal of the American Planning Association*, 63(1):107–126.
- Ewing, R. and Hamidi, S. (2015). Compactness versus sprawl. *Journal of Planning Literature*, 30(4):413–432.
- Floater, G., Rode, P., Robert, A., Kennedy, C., Hoornweg, D., Slavcheva, R., and Godfrey, N. (2014). Cities and the new climate economy: the transformative role of global urban growth. *Scientific Reports*.
- Gordon, P. and Richardson, H. W. (1997). Are compact cities a desirable planning goal? *Journal of the American Planning Association*, 63(1):95–106.
- Grover, A. and Leskovec, J. (2016). Node2vec: Scalable feature learning for networks.
- Hagberg, A. A., Schult, D. A., and Swart, P. J. (2008). Exploring network structure, dynamics, and function using networkx. In Varoquaux, G., Vaught, T., and Millman, J., editors, *Proceedings of the 7th Python in Science Conference*, pages 11 – 15, Pasadena, CA USA.
- Hamilton, W. L., Ying, R., and Leskovec, J. (2017). Inductive representation learning on large graphs.
- Hu, M.-B., Jiang, R., Wu, Y.-H., Wang, W.-X., and Wu, Q.-S. (2008). Urban traffic from the perspective of dual graph. *The European Physical Journal B*, 63.
- Jenelius, E. and Mattsson, L.-G. (2015). Road network vulnerability analysis: Conceptualization, implementation and application. *Computers, Environment and Urban Systems*, 49.
- Jepsen, T. S., Jensen, C. S., and Nielsen, T. D. (2022). Relational fusion networks: Graph convolutional networks for road networks. *IEEE Transactions on Intelligent Transportation Systems*, 23(1):418–429.
- Jepsen, T. S., Jensen, C. S., Nielsen, T. D., and Torp, K. (2018). On network embedding for machine learning on road networks: A case study on the danish road network. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 3422–3431.
- Kipf, T. N. and Welling, M. (2016). Semi-supervised classification with graph convolutional networks.
- Latora, V. and Marchiori, M. (2001). Efficient behavior of small-world networks. *Physical Review Letters*, 87(19).

- Li, X. and Liu, X. (2020). Research on identification method of key road sections in the road network under disaster situation. In *2020 5th International Conference on Electromechanical Control Technology and Transportation (ICECTT)*, pages 575–578.
- Masucci, A. P. and Molinero, C. (2016). Robustness and closeness centrality for self-organized and planned cities. *The European Physical Journal B*, 89.
- Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space.
- Mouratidis, K. (2017). Is compact city livable? the impact of compact versus sprawled neighbourhoods on neighbourhood satisfaction. *Urban Studies*, 55(11):2408–2430.
- Neuman, M. (2005). The compact city fallacy. *Journal of Planning Education and Research*, 25(1):11–26.
- OpenStreetMap contributors (2017). Planet dump retrieved from <https://planet.osm.org> . <https://www.openstreetmap.org>.
- Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab. Previous number = SIDL-WP-1999-0120.
- Perozzi, B., Al-Rfou, R., and Skiena, S. (2014). DeepWalk: Online learning of social representations.
- Pimentel, T., Castro, R., Veloso, A., and Ziviani, N. (2019). Efficient estimation of node representations in large graphs using linear contexts. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.
- Porta, S., Crucitti, P., and Latora, V. (2004). The network analysis of urban streets: A dual approach.
- Porta, S., Crucitti, P., and Latora, V. (2005). The network analysis of urban streets: A primal approach.
- Ribeiro, L. F. R., Savarese, P. H. P., and Figueiredo, D. R. (2017). Struc2vec: Learning node representations from structural identity.
- Rui, Y., Ban, Y., Wang, J., and Haas, J. (2013). Exploring the patterns and evolution of self-organized urban street networks through modeling. *The European Physical Journal B*, 86.
- Scellato, S., Cardillo, A., Latora, V., and Porta, S. (2006). The backbone of a city. *The European Physical Journal B*, 50.
- Strano, E., Nicosia, V., Latora, V., Porta, S., and Barthélemy, M. (2012). Elementary processes governing the evolution of road networks. *Scientific Reports*, 2(1).
- Uhl, J. H., Leyk, S., Chiang, Y.-Y., and Knoblock, C. A. (2022). Towards the automated large-scale reconstruction of past road networks from historical maps. *Computers, Environment and Urban Systems*, 94.
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., and Bengio, Y. (2017). Graph attention networks.
- Wilson, R. J. (1996). *Introduction to graph theory*. Prentice-Hall, London, England, 4 edition.

Xie, F. and Levinson, D. (2007). Measuring the structure of road networks. *Geographical Analysis*, 39(3):336–356.

Xu, K., Hu, W., Leskovec, J., and Jegelka, S. (2018). How powerful are graph neural networks?